The invention claimed is:

1. A method for generating photorealistic talking heads, comprising the steps of:

receiving an input stimulus;

reading data from a first library comprising images of phoneme sequences which correspond to the input stimulus;

reading, based on the data read from the first library, corresponding data from a second library comprising images of a talking subject; and

generating, using the data read from the second library, an animated sequence of a talking head tracking the input stimulus.

2. The method of claim 1, further comprising the steps of:

reading acoustic data from the second library associated with the corresponding data read from the second library;

converting the acoustic data into sound; and

outputting the sound in synchrony with the animated sequence of the talking head.

3. The method of claim 1, wherein the data read from the first library comprises parameters describing mouth shapes.

4. The method of claim 1, wherein the data read from the first library comprises one or more equations characterizing mouth shapes.

5. The method of claim 2, wherein the data read from the first library comprises one or more equations characterizing mouth shapes.

6. The method of claim 2, wherein said converting step is performed using a data-to-voice converter.

7. The method of claim 2, wherein the data read from the first library comprises segments of sampled images of a talking subject.

8. The method of claim 2, wherein the data read from the second library comprises mouth parameters characterizing degree of lip opening.

9. The method of claim 2, wherein said receiving, said generating, said converting, and all said reading steps are performed on a personal computer.

10. The method of claim 2, wherein said first and second libraries reside in a memory device on a computer.

11. The method of claim 7, wherein said first library comprises an animation library, and wherein said second library comprises a coarticulation library.

12. The method of claim 7, wherein said generating step is performed by overlaying the segments onto a common interface to create frames comprising the animated sequence.

13. A method for generating a photo-realistic talking head for a text-to-speech synthesis application, comprising the steps of:

sampling images cf a subject;

extracting a plurality of parameters from each image sample;

storing the image sample parameters into an animation library;

sampling multiphone images of the subject;

sampling sounds associated with the multiphone images;

extracting a plurality of parameters from each multiphone image sample;

storing the multiphone image parameters and associated sound samples into a coarticulation library;

reading, based on an input stimulus comprising one or more phoneme sequences, parameters from the coarticulation library corresponding to each phoneme sequence;

generating, using parameters from the animation library corresponding to the read parameters, a sequence of animated frames, the sequence tracking the input stimulus.

14. The method of claim 13, wherein the plurality of parameters extracted from each multiphone image sample comprises data describing mouth shapes.

15. The method of claim 13, wherein the plurality of parameters extracted from each multiphone image samples comprises one or more rules characterizing mouth shapes.

16. The method of claim 13, further comprising the step of:

timestamping the multiphone image samples and sound samples.

17. The method of claim 13, wherein the sound samples comprise samples converted from sound into data by a speech recognizer.

18. The method of claim 13, wherein the sound samples comprise samples converted from sound into data by a speech recognizer.

19. The method of claim 16, wherein the sound samples further comprise a phoneme transcript.

20. The method of claim 13, wherein said step of sampling images of the subject is performed by a video camera.

21. The method of claim 16, wherein said step of sampling images of the subject is performed by a video camera.

22. The method of claim 13, wherein at least one of the sampled multiphone images comprises a diphone image.

23. The method of claim 19, wherein at least one of the sampled multiphone images comprises a diphone image.

24. The method of claim 13, wherein said method is performed on a personal computer.

25. The method of claim 21, wherein said method is performed on a personal computer.

26. A processor-based method for generating a photo-realistic talking head for a text-to-speech synthesis application, comprising the steps of:

sampling images of a subject;

decomposing the subject images into a hierarchy of segments;

writing for each segment a set of parameters into memory, the segment parameter sets characterizing each segment;

sampling a plurality of phoneme sequences;

writing for each phoneme sequence a set of parameters into memory, the phoneme sequence parameter sets characterizing each phoneme sequence;

reading from memory, based upon an input stimulus, specific phoneme sequence parameter sets corresponding to the stimulus;

reading from memory, based upon the specific phoneme sequence parameter sets, corresponding specific segment parameter sets; and

generating a concatenated sequence of animated frames using the corresponding specific segment parameter sets.

27. The method of claim 26, wherein said generating step is performed by overlaying onto a common interface, for each animated frame, a plurality of segments corresponding to the specific segment parameter sets.

28. The method of claim 26, wherein said generating step comprises outputting the concatenated sequence to a screen.

29. The method of claim 27, wherein said generating step further comprises outputting the concatenated sequence to a screen.

30. The method of claim 27, wherein the segments comprise facial parts.

31. A method for generating a photo-realistic talking head for a text-to-speech synthesis application, comprising the steps of:

sampling images of a talking head;

extracting a plurality of parameters from each image sample;

writing the image sample parameters into an animation library;

sampling multiphone images of the subject;

sampling sounds associated with the multiphone images;

converting the sound samples into digital acoustic parameters;

extracting a plurality of parameters from each multiphone image sample;

storing the multiphone image parameters and associated acoustic parameters into a coarticulation library;

reading, based on an input stimulus comprising one or more phoneme sequences, parameters from the coarticulation library associated with each phoneme sequence;

generating, using parameters from the animation library, a sequence of animated frames corresponding to the read parameters and a sequence of associated sounds in synchrony with the animated frames sequence, the sequence of animated frames tracking the input stimulus.

32. The method of claim 31, wherein said converting step is performed by a speech recognizer.